

RD-R185 844

REGULARIZATION THEORY AND SHAPE CONSTRAINT(U)  
MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL  
INTELLIGENCE LAB T POGGIO ET AL. SEP 86 RI-M-916

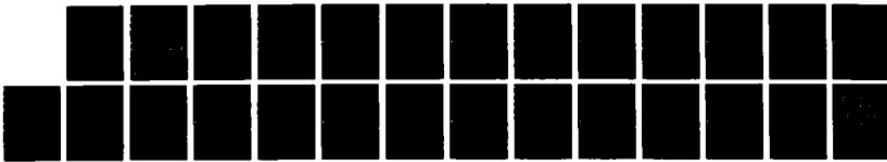
1/2

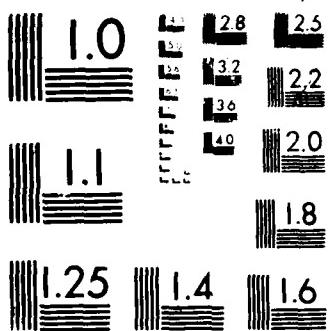
UNCLASSIFIED

DACA76-85-C-0010

F/G 12/9

ML





MICROFILM RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS 1964

AD-A185 844

DTIC FILE COPY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER MIT AIM 916	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER MIT A I Memo 916
4. TITLE (and Subtitle) "Regularization theory and shape constraint"		5. TYPE OF REPORT & PERIOD COVERED A.I. Memo
6. AUTHOR(s) Tomaso Poggio and Alessandro Verri		7. PERFORMING ORG. REPORT NUMBER DARPA DACA76-85-C-0010 DARPA/ONR N00014-85-K-0124
8. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, MA 02139		9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE September 1986
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		13. NUMBER OF PAGES 22 + cover
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		16a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Distribution is unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)  DTIC ELECTED NOV 06 1987 SD		
18. SUPPLEMENTARY NOTES  None		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Regularization Early Vision Constraints Mathematical Programming		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  → Many problems of early vision are ill-posed; to recover unique stable solutions regularization techniques can be used. These techniques lead to meaningful results, provided that solutions belong to suitable compact sets. Often some additional constraints on the shape or the behavior of the possible solutions are available. This note discusses which of these constraints can be embedded in the classic theory of regularization and how, in order to improve the quality of the recovered solution. Connections with		

Q

ABSTRACT (Continued)

mathematical programming techniques are also discussed. As a conclusion, regularization of early vision problems may be improved by the use of some constraints on the shape of the solution (such as monotonicity and upper and lower bounds), when available.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
ARTIFICIAL INTELLIGENCE LABORATORY  
and  
CENTER FOR BIOLOGICAL INFORMATION PROCESSING  
WHITAKER COLLEGE

A.I. Memo No. 916

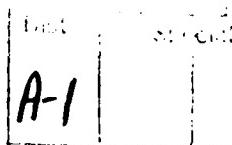
September 1986

A. Verri and T. Poggio

**Regularization Theory and Shape Constraints**

**Abstract** Many problems of early vision are ill-posed; to recover unique stable solutions regularization techniques can be used. These techniques lead to meaningful results, provided that solutions belong to suitable compact sets. Often some additional constraints on the shape or the behavior of the possible solutions are available. This note discusses which of these constraints can be embedded in the classic theory of regularization and how, in order to improve the quality of the recovered solution. Connections with mathematical programming techniques are also discussed. As a conclusion, regularization of early vision problems may be improved by the use of some constraints on the shape of the solution (such as monotonicity and upper and lower bounds), when available.

© Massachusetts Institute of Technology (1986)



This report describes research done at the Massachusetts Institute of Technology within the Artificial Intelligence Laboratory and the Center for Biological Information Processing (Whitaker College). Support for the Laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124. The Center's support is provided by a grant of the Psychology and Engineering Division of ONR and by Sloan Foundation. AV was supported by a Fairchild fellowship.

## 1. Introduction

As pointed out by Torre and Poggio (1984) many problems of early vision are ill-posed: unique stable solutions can be recovered by several regularization techniques, in particular by *standard regularization* due mainly to Tikhonov (1943, 1963). Providing that solutions belong to suitable compact sets, these techniques can be successfully applied to a broad class of problems (for a brief review see Poggio, Torre and Koch, 1985), such as surface interpolation (Grimson 1981, 1982; Terzopoulos 1984), computation of visual motion (Horn and Shunck 1981, Hildreth 1984), recovering shape from shading (Ikeuchi and Horn 1981), lightness (Horn 1974) and edge detection (Torre and Poggio 1986).

According to standard regularization theory, stable solutions can be recovered quite simply if they belong to a compact set. Otherwise standard regularization techniques have to be applied. These methods search for a solution as close as possible to the data and belonging to a compact set defined by a suitable stabilizing functional. In both cases, as we will see in detail, the concept of compact set plays a key role. Very often, however, some additional constraints on the *shape* of the possible solutions are available: for example the solutions may belong to the set of positive functions, as in the case of lightness, or may be bounded by the values of some known functions or may be piece-wise continuous or piece-wise constant as in some instances of surface interpolation. More generally it can be said that these constraints define a certain subset in a suitable functional space. Rutman and Cabral (1974) have shown that combining regularization techniques and shape constraints improves the correctness of the numerical solution in linear integral problems.

In this note, after a brief review of ill-posedness in functional spaces and in  $\Re^n$ , we show which of these constraints can be embedded in the classic regularization theory, and how. Two cases are considered in detail. In the first one, shape constraints, forcing the solution to belong to a compact set, allow a straightforward regularization of the problem. In the second, more general case, shape constraints define closed sets that can be incorporated into



the framework of classical regularization theory, where an appropriate stabilizing functional constrains the solution to a compact set, providing a simple way in which some *a priori* knowledge can be taken into account. Some functional subsets corresponding to interesting shape constraints are considered.

We also answer questions arising in the numerical solutions of regularized problems. Since regularization with shape constraints is a problem of constrained minimization, we discuss in some detail the relationship with mathematical programming.

Our main conclusion is that shape constraints can be applied in regularization theory provided they define compact or at least closed subsets. The constraints involving discontinuities do not fit into this schema while, for example, monotonicity, convexity and positivity constraints do.



## 2. Overview: ill-posed problems in infinite and finite dimensional spaces

In this section we review briefly the main problems involved in the ill-posedness of equations in infinite and finite dimensional spaces. We introduce the concepts of normal solution and quasi-solution and show the connection with uniqueness and existence of the solution to a given problem. Relationships between ill-conditioned and ill-posed problems in the discrete case are also examined.

### 2.1. Ill-posed problems in Hilbert spaces

Let us consider the problem of solving the equation

$$Ax = y \quad (2.1.1)$$



for  $x$ , where  $x$  and  $y$  belong to  $X$  and  $Y$ , Hilbert spaces. The operator  $A$ , defined on  $D(A) \subseteq X$ , maps  $D(A)$  onto  $R(A) \subseteq Y$ . In many applications it is required that the solution to (2.1.1) *i*) exists, *ii*) is unique and *iii*) depends continuously on  $y$ . A problem, whose solution satisfies *i*), *ii*) and *iii*) is said to be *well-posed* (Hadamard, 1923); otherwise it is said to be *ill-posed*. Notice that *iii*) may depend on the choice of the metric in  $X$  and in  $Y$ .

If  $A$  is linear, continuous, injective and  $R(A) = Y$ , the problem of solving (2.1.1) for  $x$  is trivially well-posed: indeed, since  $A$  is a bijection between  $D(A)$  and  $Y$ , existence and uniqueness of the solution are guaranteed. Moreover  $x$  depends continuously on  $y$  because, when  $R(A) = Y$ ,  $A^{-1}$  is continuous (Riesz and Nagy, 1952).

If  $A$  is linear but not injective, the solution to the problem of (2.1.1) is no longer unique. Uniqueness of the solution can be easily recovered, for instance, by introducing the concept of normal solution. The *normal solution*  $x_n$  to (2.1.1) is the solution orthogonal to the null space of  $A$ ,  $N(A)$ . It is easy to see that  $x_n$  is unique and that it can be characterized as the minimum norm solution. If  $A$  is injective, the normal solution and the usual solution coincide.

If we relax the condition  $R(A) = Y$  other problems arise. The solution to (2.1.1) may no longer exist since  $y$  may not belong to  $R(A)$ . For example the data  $y$  may be affected by an error  $\delta y$  belonging to the orthogonal complement to the range of  $A$ ,  $R(A)^\perp$ . It is useful then, to introduce the concept of quasi-solution (see, for example, Tikhonov and Arsenin, 1977). Let  $P$  be an operator that projects  $Y$  onto  $R(A)$ , then  $\bar{x}$ , the solution to the equation

$$Ax = Py \quad (2.1.2)$$

is called a *quasi-solution* of (2.1.1). It is obvious that  $\bar{x}$  exists if  $y \in R(A) \oplus R(A)^\perp$ . Notice that if  $y \in R(A)$ , the quasi-solution and the solution to (2.1.1) coincide.

Therefore if  $A$  is linear, continuous and  $R(A)$  is closed, the problem of finding a normal quasi-solution to the equation (2.1.1) is well-posed, since the normal quasi-solution always

exists, is unique and depends continuously on  $y$ . (This last condition follows directly from the continuity of the quasi-inverse  $A^+$  of  $A$ ,  $A^+$  being defined as the operator that maps  $y \in R(A) \oplus R(A)^\perp$  into the corresponding normal quasi-solution of (2.1.2).)

In many practical cases, however,  $R(A)$  is not closed (Kolmogorov and Fomine, 1980). So even the quasi-solution may not exist and if it exists can be unstable. Consider, for example, the Fredholm integral equation of the first kind

$$\int_a^b K(t,s)x(s)ds = y(t) \quad c \leq t \leq d. \quad (2.1.3)$$

The function

$$\tilde{x}(s) = x(s) + N \sin \omega s$$

is a solution to (2.1.3) with

$$\tilde{y}(t) = y(t) + N \int_a^b K(t,s) \sin(\omega s) ds.$$

In the usual  $L_2$  metric  $\|\tilde{y} - y\| \rightarrow 0$  as  $\omega \rightarrow \infty$  (for the Riemann Lebesgue theorem) while  $\|\tilde{x} - x\| \sim N$ . So with a suitable choice of  $N$  and  $\omega$  the error on the data can be made arbitrarily small, while the distance between the solution can be arbitrarily large.

## 2.2. Ill-posed problems in $\Re^n$

Let us consider the system of equations

$$A\mathbf{x} = \mathbf{y} \quad (2.2.1)$$

where  $A$  is a  $n \times n$  matrix and  $\mathbf{x}$  and  $\mathbf{y}$  vectors belonging to  $\Re^n$ . The problem of recovering  $\mathbf{x}$  given  $A$  and  $\mathbf{y}$  is that of finding the inverse matrix  $A^{-1}$  of  $A$ . If the determinant of  $A$  is equal to zero, the problem has no solution and the system is called *singular*. If  $A$  is diagonalizable and some eigenvalues are much smaller than the others, the system is said to be *ill-conditioned* (Strang 1976), since small errors in the data  $\mathbf{y}$  lead to unacceptable indeterminacy in the components of the solution  $\mathbf{x}$ . In such cases the ratio between the largest and the smallest eigenvalue of  $A$  is taken as the *ill-conditioning number*, that is a measure of how much the system is ill-conditioned. Notice that whether an ill-conditioning number leads to negligible errors or not depends not only on the system but also on the accuracy required.

Even in the case of huge ill-conditioning number, however, the problem of solving (2.2.1) is not ill-posed in a classical sense, since for arbitrarily small errors in the data, the solution is arbitrarily close to the exact solution. In practice, however, approximations involved in numerical computations lead to meaningless solutions, because the error in the data is not *arbitrarily small*.

Let us consider now, more closely, the problems that could arise in numerical computations: let  $\lambda_i$ ,  $i = 1, \dots, n$  be the eigenvalues of  $A$ . It is easy to see that

$$x_i = \frac{1}{\lambda_i} y_i \quad i = 1, \dots, n$$

will be the components of  $\mathbf{x}$ , the solution of (2.2.1), after a suitable transformation of coordinates. If even small errors affect the entries of  $A$ , when some  $\lambda_i$  are sufficiently close to

zero, the corresponding components  $x_i$  of the solution can become arbitrarily large, leading to an unbounded solution. As a matter of fact, the errors arising from numerical computer approximations could be sufficient; therefore even numerical problems can be ill-posed.

### 3. Shape constraints in regularization

Ill-posed problems can be successfully turned into well-posed problems by means of very general regularizing techniques. As it is well known (Tikhonov 1943, 1963), these techniques rely on the assumption of some smoothness property of the possible solution. Sometimes, however, additional and useful constraints are available; for example the solution function may be necessarily non-negative or a monotonic function and so on. In this chapter, after discussing the role of compactness in regularization, we show which of these constraints can be embedded in the classical regularization theory and how.

#### 3.1. Role of compactness in regularization

The role played by *compactness* (see Appendix A for its various definitions and properties) in the solution of ill-posed problems was clarified by Tikhonov with the following fundamental topological Lemma (Tikhonov and Arsenin, 1977):

*Lemma 3.1.1* Suppose that the operator  $A$  maps a compact set  $F \subseteq X$  onto the set  $U \subseteq Y$ ,  $X$  and  $Y$  metric spaces. If  $A : F \rightarrow U$  is continuous and one-to-one, then the inverse mapping  $A|_U^{-1}$  is also continuous.

By means of this Lemma, if the solution to equation (2.1.1) is known to belong to a compact subset of  $X$ , say  $F$ , and if the perturbed data is known to belong to  $U$ ,  $U = \{u \in Y : u = Ax\}$ , then the problem of finding a solution to (2.1.1) is trivially well-posed with respect to  $F$  and  $U$ . In such a case the problem is said to be *well-posed in the sense of Tikhonov*.

*Remark:* The compactness requirement is a strong constraint on the set of possible solutions to a given problem: it is possible to produce examples in which well-posedness is guaranteed without any compactness requirement (Groetsch, 1984).

If some *a priori* constraints on the *shape* of the solution are known and if these constraints lead to the definition of a suitable compact set, the application of Lemma 3.1.1 is straightforward. This is the theme of the next sections.

### 3.2. The selection method

A useful method of finding an approximate solution to equation (2.1.1) is the *selection method* (Tikhonov and Arsenin, 1977). It consists in calculating the operator  $A$  for points belonging to a given sample set, looking for the minimum of  $\|Ax - y\|$  in a suitable norm. Such a method is powerful from a computational point of view since the sample set can be chosen so to depend only on a finite number  $n$  of parameters varying in finite limits. Obviously the computed solution  $x_n$  and the exact solution  $x_t$  (if  $x_t$  exists) coincide if and only if  $x_t$  belongs to the sample set.

Suppose that increasing the number  $n$  of parameters (and therefore the dimension of the subspace containing the sample set)  $\|Ax_n - y\| \rightarrow 0$ . Let us assume, therefore that  $\|Ax_n - y\| \rightarrow 0$  as  $n \rightarrow \infty$ . It is easy to see that if the  $R(A)$  is not closed the approximate solution  $\|x_n\| \rightarrow \infty$ , hence  $x_n$  does not converge to  $x_t$ . In order to guarantee the convergence of  $x_n$  to  $x_t$ , compactness of the sample set is needed, so that Lemma 3.1.1 applies. If the sample set is not compact but it is closed and bounded, the Lemma 3.1.1 is still valid, though in a weaker sense. The solution  $x_n$ , in fact, is only weakly convergent<sup>1</sup> to the true solution  $x_t$ : it is also convergent in the usual sense if  $x_t$  lies on the boundary of the sample set (Bertero, 1982).

---

<sup>1</sup> The solution  $x_n$  is said to be weakly convergent to  $x_t$  if  $(x_n, y) \rightarrow (x_t, y)$  for  $n \rightarrow \infty \forall y \in X$ , where  $(\cdot)$  is a suitable dot product.

### 3.3. Regularization theory and shape constraints

When no compact set containing the possible solution of (2.1.1) can be found, a new approach is needed. A general and useful approach was also outlined by Tikhonov (1943, 1963) and is called standard *regularization theory*. Let us briefly summarize the main points of this theory.

The fundamental concept of the theory is that of a regularizing operator. Suppose that the equation (2.1.1) allows  $x = x_t$  as a solution when  $y = y_t$ ; then an operator  $R(y, \alpha)$  is called a *regularizing operator* for the equation (2.1.1) in a neighborhood of  $x = x_t$  if:

- i)  $\exists \delta_1 > 0$  such that  $R(y, \alpha)$  is defined  $\forall \alpha > 0$  and  $\forall y \in Y$  such that  $\|y - y_t\| \leq \delta_1$ :
- ii) there exists a function  $\alpha = \alpha(\delta)$  such that  $\forall \epsilon > 0 \exists \delta \leq \delta_1$  such that  $\forall y$

$$\|y - y_t\| \leq \delta \implies \|x_t - x_\alpha\| \leq \epsilon$$

where  $x_\alpha = R(y, \alpha(\delta))$

So the problem of finding a regularized solution to an ill-posed problem is shifted to that of finding methods to construct a regularizing operator. Let us see in some detail one of these methods.

#### *Construction of regularizing operators by minimization of a smoothing functional*

It is possible to construct a regularizing operator for (2.1.1) by minimizing the following functional with respect to  $x$ :

$$\Psi^\alpha[x, y] = \|Ax - y\| + \alpha \Omega[x] \quad (3.3.1)$$

where  $\Omega$  is a *stabilizing functional*. A functional  $\Omega$  defined on  $O \subset D(A)$  everywhere dense in  $D(A)$  is a stabilizing functional for the equation (2.1.1) if:

- i)  $x_t$  belongs to the domain of definition of  $\Omega$ ;
- ii)  $\forall d > 0$ ,  $\{x \in O \mid \Omega[x] \leq d\}$  is a compact subset of  $O$ .

Indeed the following theorem holds:

*Theorem 3.3.1* Let  $A$  denote a continuous operator. For every  $y \in Y$  and every  $\alpha > 0$ , there exists a  $x_\alpha \in O$  for which the functional  $\Psi$  attains its minimum.

As a matter of fact the choice of  $\Omega$  can determine the uniqueness of the solution: for example if  $D(A)$  is a Hilbert space and  $A$  is linear, if  $\Omega$  is quadratic, sufficient condition for the uniqueness of the regularized solution can be proved (Tikhonov and Arsenin, 1977). In principle, the regularization problem is completely solved. Sometimes, however, some additional constraints on the shape of the solution are available. Can we exploit them? Indeed, the following Lemma holds:

*Lemma 3.3.2* Let  $X$  be a compact topological space. Then every closed subset of  $X$  is compact.

Theorem 3.3.1 is based on the compactness of the subsets where  $\Omega$  is bounded and therefore is still true even if the set of possible solutions is a closed subset of  $D(A)$ . Therefore, if the additional constraints lead to the definition of some closed subset of  $D(A)$ , they can be easily exploited in the framework of regularization theory.

*Remark:* these sets do not need to be compact. The regularizing scheme itself provides compactness of the set in which the solution is actually searched: if the constraints define a compact set, the Lemma 3.1.1 is sufficient to guarantee well-posedness of the problem.

### 3.4. Compact subsets of functional spaces

From the preceding sections, it turns out that given an ill-posed problem and some *a priori* constraints, it is important to determine whether such constraints define a compact subset or at least a closed subset of a suitable functional space. Let us examine some examples of subsets of  $L_2$  and  $C^0$ .

The set of bounded non-decreasing (non-increasing) functions is a compact set. The proof (see Taylor, 1965, for example) relies upon the fact that the number of discontinuity points of a monotonic bounded function is at most enumerable.

The set of convex functions is compact. This result follows trivially from the compactness of the set above, since each convex function is the integral of a suitable non-decreasing function.

The set of bounded piece-wise constant functions is neither closed nor compact. It is not compact since it is everywhere dense in  $L_2$  (which is trivially not compact). It is not closed since any continuous function is an accumulation point of this set.

It is not easy to find compact subsets of  $C^0$ . The set of bounded non-negative functions, for example, is not compact. Consider in  $C[0, 1]$

$$T = \{x \mid |x(t)| \leq 1, t \in [0, 1]\}.$$

$T$  is closed and bounded (obvious), but not compact. Indeed, let  $S = \{x_i\}_{i \in N}$  be a sequence of functions with  $x_n(t) = t^n$ . Any subsequence of  $S$  cannot converge in  $T$ , since in  $C[0, 1]$  the convergence is uniform convergence while  $t^n \rightarrow 0$ , if  $0 \leq t < 1$ , and  $t^n = 1$ , if  $t = 1$ . So  $T$  is not compact.

*Remark:* This counterexample shows that in  $C^0$  even the sets of monotonic and convex functions are not compact.

As a conclusion, the constraints of monotonicity and of convexity, defining compact subsets, can be useful in regularization either via the selection method or via standard techniques (since any compact set is closed, see Appendix A for detail). The positivity constraint can be used only as a shape constraint in classical regularization theory and in a weaker sense in the selection method, while piece-wise constant functions, though representing significant *a priori* knowledge on the shape of the solution, cannot be embedded in either of the frameworks.

## 4. Connection with mathematical programming (MP)

Most of the problems faced in the framework of Hilbert spaces are in fact usually either intrinsically discrete problems or problems allowing only numerical solution. In this chapter the cases of the selection method and of the regularization with shape constraint, discussed in the preceding sections, are analyzed in this respect as examples of mathematical programming problems.

### 4.1. Selection method as a MP problem

As we have seen in the previous chapter, if the condition of section 3.1 applies, an approximate solution to the equation (2.1.1) can be found by means of the selection method. In practice the problem has to be solved numerically: consider for example the Fredholm integral equation of the first kind

$$\int_a^b K(t,s)x(s)ds = y(t) \quad c \leq t \leq d \quad (4.1.1)$$

where  $x(s)$  belongs to a set  $F$  of decreasing uniformly bounded functions.  $F$  is compact (see section 3.4), therefore if  $y(t) \in U = AF$  the problem is *well-posed in the sense of Tikhonov*. In order to find an explicit solution we can replace the integral with a sum over a grid with  $n$  nodes. Let  $x_i$  ( $i = 1, \dots, n$ ) be the value of the unknown vector  $\mathbf{x}$  at the node  $i$  and  $y_j$  ( $j = 1, \dots, m$ ) the components of the data vector  $\mathbf{y}$ . The problem is to find a bounded vector minimizing the functional

$$\Psi[\mathbf{x}, \mathbf{y}] = \left\| \sum_{j=1}^m \sum_{i=1}^n (K_{ji}x_i - y_j) \right\|$$

under the constraint that the components of  $\mathbf{x}$  are decreasing. It is easy to show that this constraint can be expressed as a positivity constraint on the values of the derivative of

the function at each node. In the discrete case this reduces to the fact that suitable linear combinations of the neighbor nodes have to be greater than zero. For example in the nearest neighbor approximation we have

$$\frac{(x_{i+1} - x_{i-1})}{2} > 0 \quad i = 2, \dots, n-1. \quad (4.1.2)$$

In these terms the problem is now a typical problem of quadratic programming (see Appendix B for main definitions and results of mathematical programming problems). Indeed in the general case the only problem concerns the explicit form of the constraints. It must be possible to write them as follows (see Appendix B):

$$g_i(\mathbf{x}) \leq 0 \quad i = 1, \dots, c \quad (4.1.3)$$

where  $g_i$  are scalar functions (they need to be linear or at most quadratic to define a quadratic programming problem). Notice that (4.1.2) can be immediately rewritten like (4.1.3). Rutman and Cabral (1974) have shown that performing a suitable transformation, the constraint of monotonicity, convexity, unimodality and selective non-negativity can all be written in the form (4.1.2). In this case, as shown before, only the monotonicity and the convexity constraint can be properly used. As we will see in the next section, however, all of them are shape constraints that can be useful in regularization.

## 4.2. Regularization with shape constraint as an MP problem

Let us illustrate this section by means of the same example of the previous one. Again the problem is to solve the Fredholm integral equation of the first kind (4.1.1). This time since either the set  $F$  is not compact or  $y$  does not belong to  $U = AF$ , standard regularization techniques of the kind described in section 3.3 are needed. Suppose moreover that some further information is available and that they correspond to constraints on the solution defining closed subsets of the domain of the operator. If these constraints can be written in

the form (4.1.2) the problem of minimizing the discrete functional corresponding to (3.3.1) subject to such constraints is again a typical stable problem of quadratic programming.

*Remark:* A generic mathematical programming problem, even if quadratic or linear, is not necessarily stable. As a matter of fact the well-posedness relies on the strong assumption that the functional to minimize is a stabilizing functional. If this is not the case, the problem has to be regularized following standard techniques (Tikhonov and Arsenin, 1977).

*Remark:* While any regularized problem of the type described in section 3.3 gives rise to a well-posed mathematical problem, the application of Kuhn-Tucker theory and of the gradient method are subject essentially to the fulfillment of some convexity properties of the functions involved (see Appendix B) and therefore they are guaranteed only in the case of linear operators and a quadratic stabilizing functional.

## 5. Conclusion

In this note we analysed the role played by shape constraints in ill-posed problems. The key concept has been that of compact set. If the shape constraints lead to the definition of a compact set, regularization is straightforward. Indeed the shape constraint itself provides sufficient conditions for the continuity of the dependence of the solution on the data. If the shape constraints define at least a closed set, then they can be an useful addition to standard regularization approaches. While a suitable functional provides stability on the data, shape constraints allow to recover a solution *closer* to the correct one, by taking into account significative additional *a priori* knowledge on the shape of the solution.

In both cases constraints that do not define at least a closed set cannot be embedded in the regularizing step. In particular this implies that the *a priori* knowledge concerning piece-wise constant or piece-wise continuous functions, though in principle significant for many early vision problems (the reconstruction of the 3D structure of a scene and the

recovery of the *albedo* for example) cannot be used within any classical regularizing schema. This is an additional argument that motivates the use of Markov Random Fields models for exploiting *a priori* information about discontinuities and their properties (see Marroquin et al. 1985). A different regularizing approach that can exploit constraints of this type, considering discrete and quantized formulations, will be discussed in a forthcoming paper (Poggio and Verri).

Finally, the discrete problem that has to be faced solving an ill-posed problem has been analysed as a mathematical programming problem: in the interesting case of linear operators it becomes a standard stable problem of quadratic programming. In particular, all the results of convex programming regarding local and global convergence of the gradient method algorithm are guaranteed to apply.

### Acknowledgement

We would like to thank M. Bertero and V. Torre for their useful suggestions and discussions. E. Hildreth and E. Grimson read the manuscript and made valuable comments.

## Appendix A: Compact sets in topological and metric spaces

As we have seen before the concept of compact set is fundamental in the regularization of ill-posed problems. Unfortunately, there exist different definitions of compact set. Disregarding historical problems, here is a summary of the main definitions and properties concerning compact sets that we adopted in this note.

Let  $X$  be a topological space. An *open covering* of  $S \subseteq X$  is a family  $\Gamma$  of open sets in  $X$  such that  $S \subseteq \bigcup_{i \in \Gamma} i$ .

$S \subseteq X$  is *compact* if, for every open covering  $\Gamma$  of  $S$ , there exists a finite subfamily of  $\Gamma$  that also covers  $S$ .

*Remark:* A closed set is not necessarily compact (consider the real line). A compact set is not necessarily closed. Compact sets are always closed in Hausdorff spaces (a topological space is a Hausdorff space if for each pair of distinct points  $x_1$  and  $x_2$ , there exist two disjoint neighborhoods containing them).

In topological spaces the following Lemma holds:

*Lemma A.1* If  $T \subseteq X$  is compact, then for every infinite  $S \subseteq T$ ,  $S' \cap T \neq \emptyset$ . ( $S'$  is the set of accumulation points of  $S$ ).

Notice that the converse of Lemma A.1 is not true in general. Now let  $X$  be a metric space (and henceforth a topological Hausdorff space) then we have:

*Lemma A.2* If  $T \subseteq X$  and for every infinite  $S \subseteq T$ ,  $S' \cap T \neq \emptyset$ , then  $T$  is compact.

*Remark:* Combining Lemma A.1 and A.2 the usual definition of *compactness* in metric spaces can be obtained: a set  $S \subseteq X$ ,  $X$  a metric space, is *compact* if for every sequence of points in  $S$  there is a subsequence converging to a point of  $S$ .

Furthermore, in metric spaces the concept of *boundedness* can be defined, so that the following Lemma can be proved:

*Lemma A.3* If  $S \subseteq X$  is compact then  $S$  is closed and bounded.

The converse of Lemma A.3 is not true in general (see section 3.4 for a counterexample).

*Remark:* In  $\mathbb{R}^n$  the converse of Lemma A.3 holds. Indeed in  $\mathbb{R}^n$ , for the Borel theorem, any bounded set has an accumulation point; so, if it is closed, it is also compact.

It follows that discretization makes a problem well-posed (but ill-conditioned possibly).

## Appendix B: Mathematical programming: Definitions and main results

In Vision, when a regularized problem has to be solved, numerical methods, based on discretizing the original continuous formulation, are usually needed. These numerical methods always lead to classical problems of mathematical programming. As may be expected in these cases Tikhonov regularization theory and mathematical programming theorems predict the same results in terms of existence and uniqueness of the solution (see sections 4.1 and 4.2). Here we review, for the sake of completeness, the main definitions and results of mathematical programming theory (for more details see Anew et al., 1958, for example).

Let us consider the problem of finding a minimum for a given functional  $\varphi = \varphi(\mathbf{z})$  on a set  $G = \{\mathbf{z} / g_i(\mathbf{z}) \leq 0, i = 1, \dots, m\}$  where  $\mathbf{z} = (z_1, \dots, z_n) \in L \subseteq \mathbb{R}^n$  and  $g_i$  are scalar functions. If the functions  $\varphi$  and  $g_i$  ( $i = 1, \dots, m$ ) are linear, the problem is called a *linear programming problem*, otherwise *non-linear*. In both cases it is a *mathematical programming problem*.

Typically the problem of finding conditional extrema of a given functional is solved by means of the Lagrange multipliers theory. Classical theorems on Lagrange multipliers provide only necessary conditions for the existence of such multipliers: Kuhn-Tucker theory, in turn, fills the gap, providing sufficient conditions for their existence (obviously closely related to the existence of extrema of functionals). This theory, therefore, is useful in most of the mathematical programming problems. Let us review briefly the main results of Kuhn-Tucker theory.

### Kuhn-Tucker theory

Let us call the conditional problem stated above *P.1* and associate with it the following Lagrangian form

$$\Phi(\mathbf{z}, \mathbf{w}) = \varphi(\mathbf{z}) + \sum_{i=1}^m w_i g_i(\mathbf{z})$$

where  $\mathbf{w} = (w_1, \dots, w_m)$  with  $w_i \in \Re^+$ ,  $i = 1, \dots, m$ . It is easy to see that if the pair  $(\mathbf{z}', \mathbf{w}')$  is a saddle point for the above Lagrangian form  $\mathbf{z}'$  is a solution of *P.1*. Let us call *P.2* the problem of finding a saddle point for the Lagrangian form  $\Phi$ , thus the following Lemma holds:

*Lemma B.1* Given *P.1* and *P.2*, if the pair  $(\mathbf{z}', \mathbf{w}')$  is a solution to *P.2* then  $\mathbf{z}'$  is a solution to *P.1*.

To prove the converse of *Lemma B.1*, i.e. to show the equivalence between *P.1* and *P.2*, some constraints on the functions  $\varphi$  and  $g_i$ ,  $i = 1, \dots, m$  are needed; more precisely:

*Theorem B.2 (Kuhn-Tucker)* Let  $\varphi(\mathbf{z})$  and  $g_i(\mathbf{z})$ ,  $i = 1, \dots, m$  be convex on  $Z = \{\mathbf{z} / z_i \geq 0, i = 1, \dots, n\}$ . If there exists  $\mathbf{z}^0 \in Z$  such that  $g_i(\mathbf{z}^0) < 0$ ,  $i = 1, \dots, m$ , then  $\mathbf{z}'$  is a solution to *P.1* if and only if  $\exists \mathbf{w}'$  such that the pair  $(\mathbf{z}', \mathbf{w}')$  is a solution to *P.2*.

In the case of  $C^1$  functions the celebrated *Kuhn-Tucker conditions* can be introduced. They guarantee necessary conditions for the existence of a solution to a saddle point problem. Under convexity assumptions the Kuhn-Tucker conditions become sufficient, henceforth guaranteeing the existence of a solution to the associated mathematical problem. (If  $\varphi$  is strictly convex it also turns out that the solution is unique). In obvious notation they are:

$$\sum_{j=1}^m w'_j \frac{\partial \Phi}{\partial w_j}(\mathbf{z}', \mathbf{w}') = 0$$

$$\frac{\partial \Phi}{\partial w_j} \leq 0, j = 1, \dots, m$$

$$w_j \geq 0, \quad j = 1, \dots, m$$

$$\sum_{i=1}^n z'_i \frac{\partial \Phi}{\partial z_i}(\mathbf{z}', \mathbf{w}') = 0$$

$$\frac{\partial \Phi}{\partial z_i} \geq 0, \quad i = 1, \dots, n$$

$$z_i \geq 0, \quad i = 1, \dots, n.$$

### *Gradient method*

Let us now review briefly the *gradient method*, which is one of the most useful methods for finding saddle points of a given function. It consists essentially in finding the solution of the following system  $S$  of differential equations

$$\frac{dz_i}{dt} = 0 \quad \text{if} \quad \frac{\partial \Phi}{\partial z_i} > 0 \quad \text{and} \quad z_i = 0$$

$$\frac{dz_i}{dt} = -\frac{\partial \Phi}{\partial z_i} \quad \text{otherwise;} \quad i = 1, \dots, n$$

$$\frac{dw_j}{dt} = 0 \quad \text{if} \quad \frac{\partial \Phi}{\partial w_j} < 0 \quad \text{and} \quad w_j = 0$$

$$\frac{dw_j}{dt} = \frac{\partial \Phi}{\partial w_j} \quad \text{otherwise;} \quad j = 1, \dots, m$$

where  $t$  is a parameter. Now if the pair  $(\mathbf{z}', \mathbf{w}')$  is a saddle point for  $\Phi(\mathbf{z}, \mathbf{w})$  it follows that:

$$\frac{\partial \Phi}{\partial z_i}(\mathbf{z}', \mathbf{w}') \leq 0 \quad i = 1, \dots, n$$

$$\frac{\partial \Phi}{\partial w_j}(\mathbf{z}', \mathbf{w}') \leq 0 \quad j = 1, \dots, m.$$

In particular if  $\frac{\partial \Phi}{\partial z_i}(\mathbf{z}', \mathbf{w}') > 0$  then  $z'_i = 0$  and if  $\frac{\partial \Phi}{\partial w_j}(\mathbf{z}', \mathbf{w}') < 0$  then  $w'_j = 0$ . Without loss of generality (just for notational convenience) suppose in the sequel that for  $i = 1, \dots, p$ ,  $\frac{\partial \Phi}{\partial z_i}(\mathbf{z}', \mathbf{w}') = 0$ , while for  $i = p+1, \dots, n$   $\frac{\partial \Phi}{\partial z_i}(\mathbf{z}', \mathbf{w}') > 0$  and that for  $j = 1, \dots, q$ ,  $\frac{\partial \Phi}{\partial w_j}(\mathbf{z}', \mathbf{w}') = 0$ , while for  $j = q+1, \dots, m$   $\frac{\partial \Phi}{\partial w_j}(\mathbf{z}', \mathbf{w}') < 0$ .

The following theorem, now, guarantees local convergence of the gradient method.

*Theorem B.3* Let  $\Phi(\mathbf{z}, \mathbf{w})$  have a saddle point  $(\mathbf{z}', \mathbf{w}')$  under the constraint  $\mathbf{z} \in Z$  where  $Z = \{\mathbf{z} / z_i \geq 0, i = 1, \dots, n\}$  and  $\mathbf{w} \in W$  where  $W = \{\mathbf{w} / w_j \geq 0, j = 1, \dots, m\}$  and let  $\Phi$  be analytic in some neighborhood of  $(\mathbf{z}', \mathbf{w}')$ . Suppose further that the matrix of the second derivative of  $\Phi$  in the first  $p$  components of  $\mathbf{z}$  defines a positively defined form and that  $z_i > 0$ ,  $i = 1, \dots, p$  and  $w_j > 0$ ,  $j = 1, \dots, q$ . Then for any pair  $(\mathbf{z}'', \mathbf{w}'')$  in a sufficiently small neighborhood of  $(\mathbf{z}', \mathbf{w}')$ :

- i) there is a unique solution  $\mathbf{z} = \mathbf{z}(t, \mathbf{z}'', \mathbf{w}'')$  and  $\mathbf{w} = \mathbf{w}(t, \mathbf{z}'', \mathbf{w}'')$  to the system  $S$  such that:
- ii)  $\lim_{t \rightarrow \infty} \mathbf{z}(t, \mathbf{z}'', \mathbf{w}'') = \mathbf{z}'$  and
- iii) in any limit point  $\mathbf{w}^0$  of the function  $\mathbf{w} = (t, \mathbf{z}'', \mathbf{w}'')$  as  $t \rightarrow \infty$ , the pair  $(\mathbf{z}', \mathbf{w}^0)$  is saddle point of  $\Phi(\mathbf{z}, \mathbf{w})$ .

*Remark:* The classical theorems of existence and uniqueness of the solution for differential system of equation cannot be used, since no assumption is actually made on the continuity of the derivatives of the variables.

Before stating the theorem on global stability of the gradient method the following definition is needed:

$z_i(t)$ ,  $i = 1, \dots, n$  and  $w_j(t)$ ,  $j = 1, \dots, m$ , solution of the system  $S$  are a *regular* solution if when  $z_i(t_\nu) = 0$ ,  $i = 1, \dots, n$  and  $w_j(t_\nu) = 0$ ,  $j = 1, \dots, m$  with  $\nu \in N$  for some sequence

$\{t_\nu\}$  such that  $t_\nu > 0$  and  $\lim_{\nu \rightarrow \infty} t_\nu = 0$ , there is some  $\tilde{t} > 0$  such that  $\frac{d\varphi}{dt} = 0$ ,  $i = 1, \dots, n$  and  $\frac{d\varphi_j}{dt} < 0$ ,  $j = 1, \dots, m$  for  $0 < t < \tilde{t}$ .

*Theorem B.4* Let  $\Phi(\mathbf{z}, \mathbf{w})$  be a strictly convex, continuous and twice differentiable function in  $\mathbf{z} \in Z$  and  $\mathbf{w} \in W$ . Let the system  $S$  have a regular solution with respect to any pair  $(\mathbf{z}'', \mathbf{w}'')$  where  $\mathbf{z}'' \in Z$  and  $\mathbf{w}'' \in W$ . Then there is a unique regular solution of the system  $S$  with any initial position. Furthermore if  $\Phi$  has a saddle point in  $(\mathbf{z}', \mathbf{w}')$  under the constraints  $\mathbf{z} \in Z$  and  $\mathbf{w} \in W$ ,  $\mathbf{z}'$  is uniquely determined and any solution of  $S$  converges to  $\mathbf{z}'$ .

*Remark:* Actually by introducing suitable strictly increasing functions  $\rho_j$ ,  $j = 1, \dots, m$  of one variable such that  $\rho_j(0) = 0$ ,  $j = 1, \dots, m$ , the condition of strict convexity in theorem B.4 can be relaxed to convexity if one applies the gradient method to the *modified Lagrangian* defined

$$\Phi_\rho(\mathbf{z}, \mathbf{w}) = \varphi(\mathbf{z}) + \sum_{j=1}^m w_j \rho_j[g_j(\mathbf{z})].$$

In conclusion Theorem B.4 guarantees global convergence of the gradient method for convex programming (including therefore the important case of quadratic programming); the modified Lagrangian form above allows the successful extension of the gradient method to the broad class of linear programming problems.

## References

- Arrow, K.J., Hurwicz, L. and Uzawa, H. **Studies in linear and non-linear programming**. Stanford University Press, Stanford, Ca., 1958.
- Bertero, M. **Problemi lineari non ben posti e metodi di regolarizzazione** in *Problemi non ben posti ed inversi*. Istituto di Analisi Globale, firenze, 1982.
- Grimson, W.E.L. **From images to surfaces**, MIT Press, Cambridge, Mass 1981.
- Grimson, W.E.L. "A computational theory of visual surface interpolation", *Phil. Trans. R Soc. London, B*, 298, 395-427, 1982
- Groetsch, C.W. **The theory of Tikhonov regularization for Fredholm integral of the first kind**, Pitman Advanced Publ. Program, 1984.
- Hadamard, J. **Lectures on Cauchy's problem in linear partial differential equations**. Yale University Press, 1923.
- Horn, B.K.P. "Determining lightness from an image" *Computer Graphics and Image Processing*, **3**, 1974.
- Horn, B.K.P. and Schunck, B.G. "Determining optical flow", *Artificial Intelligence*, **17**, 155-203, 1981.
- Hildreth, E.C. "Computation of the velocity field", *Proc. R. Soc. London, B*, 221, 189-220, 1984.
- Ikeuchi, K. and Horn, B.K.P. "Numerical shape from shading and occluding boundaries". *Artificial Intelligence*, **17**, 141-184, 1981.
- Kolmogorov, A. and Fomine, S. **Elements de la Theorie des function et de l' analyse fonctionnelle** Edition M.I.R., Moscow, 1980.
- Marroquin, J., Mitter, S. and Poggio, T. "Probabilistic Solution of Ill-posed Problems in Computational Vision", *Proceedings of the Image Understanding Workshop*, Miami Beach, 1985.

- Poggio, T., Torre, V. and Koch, K. "Computational Vision and Regularization Theory", *Nature* **317**, 1985.
- Poggio, T. and Verri, A. "The constraint method: a regularization approach to early vision problems" (in preparation).
- Turcsanyi, F. and Nagy, B. **Lecons d' analyse fonctionnelle**. Budapest, 1952.
- Rudin, L., R.S. and Cabral, L.M. "Descriptive regularization of the Fredholm integral equations of the first kind"
- Stern, G. **Introduction to linear algebra and its applications**. Academic Press, 1976.
- Halmos, P.R. **General theory of functions and integration**. Blaisdell Publ. Company, New York, 1965.
- Bergeropoulos, D. "Multiresolution computation of visible-surface representation", Ph.D. Thesis, MIT, 1984.
- Tikhonov, A.N. "The stability of inverse problems" *Doklady Akad. Nauk SSSR*, **39**, 5, 1943.
- Tikhonov, A.N. "Regularization of ill-posed problems" *Doklady Akad. Nauk SSSR*, **153**, 1, 1963.
- Tikhonov, A.N. and Arsenin, V.Y. **Solutions of ill-posed problems**. W.H.Winston, Washington, D.C., 1977.
- Hochman, E.C. **Eigenfunction expansions associated with second order differential equations**, part II, Oxford University Press, 1958.
- Torre, V. and Poggio, T. "Ill-posed problems and regularization analysis in early vision" - Artificial Intelligence Lab. Memo No 723, MIT Cambridge, Ma, 1984a
- Torre, V. and Poggio, T. "On edge detection" *IEEE transactions on Pattern Analysis and Machine Intelligence*, **8**, No 2, 1986.

END  
DATE  
FILMED  
JAN  
1988